Determinants of Intraregional Migration in Sub-Saharan Africa 1980-2000

Ilse Ruyssen and Glenn Rayp*

SHERPPA, Ghent University

August 2013

Abstract

Despite great accomplishments in the migration literature, the determinants of South-South migration remain poorly understood. In an attempt to fill this gap, this chapter formulates and tests an empirical model for intraregional migration in Sub-Saharan Africa within an extended human capital framework, taking into account spatial interaction. Using bilateral panel data between 1980-2000, we find that intraregional migration on the subcontinent is predominantly driven by economic opportunities and sociopolitics in the host country, facilitated by geographical proximity. The role played by network effects and environmental conditions is also apparent. Finally, origin and destination spatial dependence should definitely not be ignored.

JEL Classification: F22, O15, C23

Keywords: International migration, Sub-Saharan Africa, Spatial interaction, Spatial Durbin model

1 Introduction

The motivations for international migration have received a great deal of attention in migration research since the 1980s. The main focus of theoretical and empirical research has been on the principal channels of mass migration in the twentieth century. These include both North-North migration, such as European migration to North America or Australia, as well as South-North migration, such as migration from former colonies to Europe and migration in the context of guest worker programs and exile. Recent empirical studies have typically analyzed migration to Europe (Gallardo-Sejas et al., 2006; Hooghe et al., 2008) or to the OECD (Pedersen et al., 2008; Ortega and Peri, 2009, 2011; Mayda, 2010; Beine et al., 2011; Ruyssen et al., 2012), estimating a variant of the human capital model of migration with particular attention to economic determinants.

The driving forces behind migration to developing countries, especially South-South migration, remain poorly understood. Yet, the extent of migration in the South should definitely not be underestimated. The

*Corresponding author: Ilse Ruyssen, Department of Economics, Ghent University, Tweekerkenstraat 2, B-9000 Gent, Belgium. Tel.: +32 9 264 42 42. Fax: +32 9 264 35 99. Email: Ilse.Ruyssen@UGent.be.

World Bank estimated that in 2000, 51 percent of worldwide migration could be classified as migration to the South[1]. This implies that in 2000, 85 million out of 165 million migrants on the globe were living in a developing country. In Sub-Saharan Africa (SSA), the extent of South-South migration goes even beyond that of South-North migration, with as much as 69 percent of the movement classified as South-South migration. The share of migration to other developing regions (interregional migration) is negligible, suggesting a great deal of intraregional migration on the African subcontinent.[2] The relatively little scholarly attention that international migration within Africa south of the Sahara has received can primarily be linked to the lack of reliable data. Despite great improvements in the availability of international migration data during recent years, detailed long-term data on immigrant flows remain unavailable or incomplete for many developing countries. This is especially the case for the relatively poorer African countries, for which keeping track of border crossings has not been a priority on the policy agenda.

Because data on international SSA migration is scarce, most of the literature dealing with migration in SSA has focused on rural-urban migratory movements within countries (Agesa and Agesa, 1999; Andersson, 2001; de Haan et al., 2002; Hampshire, 2002). Barkley and McMillan (1994), for instance, estimated a migration decision model incorporating both economic conditions as well as political institutions, using panel World Bank data for 32 African countries during 1972-1987. They found support for their hypothesis that the presence of political freedom and civil liberties augments the responsiveness of labor migration to economic incentives. Alternatively, Barrios et al. (2006) analyzed the impact of environmental change on urbanization in SSA using a panel of 78 countries between 1960-1990. They confirmed that, contrary to the results for other developing regions, shortages in rainfall have acted to increase rural-urban movements in SSA countries.

Studies that analyze intraregional SSA migration, on the other hand, typically focus on migration to the south and the west, and mainly involve case studies such as mine migration in South-Africa (Lucas, 1985, 1987; Taylor, 1990), war-related border crossing between Zimbabwe and Mozambique (Hughes, 1999) or Mozambican refugee flows to Malawi (Koser, 1997). To our knowledge, only a few studies have tried to empirically investigate the determinants of intraregional SSA migration on a more comprehensive level. Hatton and Williamson (2002), for instance, estimated the determinants of net out-migration rates (calculated as a residual from demographic accounting) in countries across SSA. They found that Africans are especially driven by wage gaps and demographic booms in the sending country. However, as the authors had no information about the migrants' origin or destination, these results only offer an indication of the motivations for emigration from developing countries, but not necessarily for South-South migration.

The recently constructed Global Bilateral Migration Database (GBMD) described in Özden et al. (2011), however, offers new opportunities to exploit bilateral panel data to investigate incentives for South-South migration as is usually done in a South-North context. Spanning the period 1960-2000, it is the most comprehensive and consistent database on bilateral South-South migration available at present.

[1]We follow Özden et al. (2011) who classify Australia, Canada, Japan, New Zealand, the United States, the EU-15 and the European Free Trade Association as developed countries, the remaining countries being classified as developing.

[2]For a detailed overview of migratory patterns in SSA see Adepoju (1995), Adebusoye (2006) and Ncube et al. (2010).

The database provides statistics on migrant stocks for each decade during this period. The change in migrant stocks between subsequent time periods can then be used as a measure of net migration flows (see also Beine et al., 2011; Marchiori et al., 2012). This approximation is not perfect as it does not take into account deaths and return migration during the 10 years between observation points. Yet, following Beine et al. (2011), we believe that it is accurate enough to provide a reasonable approximation for net migration.

As such, the first contribution of this chapter concerns the use of bilateral panel data to evaluate the factors affecting migration between SSA countries. We specify a comprehensive human capital model of migration that encompasses not only the typical economic determinants of migration but also demographic, sociopolitical and environmental factors representing characteristics of countries of origin and destination as well as network effects and natural and cultural factors enhancing or restraining migrant flows to the host country, such as transport, communication and psychological costs of migration. The model is estimated using data from the GBMD, for 42 origin and destination countries between 1980-1990 and 1990-2000.

The second contribution of this chapter relates to our estimation approach, which takes into account potential spatial interaction between origin-destination (OD) flows. As argued by Griffith and Jones (1980), OD flows from a certain origin (to a certain destination) are positively correlated with the degree of emissiveness (attractiveness) of its neighboring origin (destination) locations. Although several authors have pointed out the need to account for spatial dependence in the analysis of migratory movements (see for example Cushing and Poot, 2003), the use of spatial regression methods in the migration literature is still limited. To address this apparent gap in the literature, LeSage and Pace (2008) develop a family of spatial OD models using a combination of three spatial connectivity matrices for destination, origin and destination-to-origin dependence which can be estimated using maximum-likelihood techniques. We follow this approach, which allows for a general structure of the spatial correlation in the migrant flow. Starting from a spatial Durbin model, the most general model of spatial dependence, we rely on specification tests to determine which model best describes the data.

Our main results can be summarized as follows. Although we find evidence for a strong influence of average incomes in the host country, the role played by sociopolitical factors is also apparent, though only indirectly. The occurrence of conflict in the home country encourages emigration towards countries where relative freedom is secured. These migratory streams are perpetuated because of network effects lowering the psychological costs of migration. Also distance and adjacency play a significant role because of their influence on transport and communication costs. It is shown that the influence of environmental factors should not be underestimated: immigration is higher towards countries with lower disaster occurrence and indirectly also temperature anomalies. Finally, we find indications of significant destination- and origin-based spatial dependence in migration decisions.

The remainder of the chapter is organized as follows. Section 2 outlines the empirical model. Section 3 describes the data. The introduction of spatial dependence and the estimation method are discussed in Section 4. Section 5 elaborates on the estimation results and Section 6 concludes.

2 Empirical model specification

As most of the recent economic literature on the migration decision (see Hatton, 1995; Pedersen et al., 2008; Mayda, 2010; Ruyssen et al., 2012), our empirical model is based on Sjaastad's (1962) human capital model of migration. Economic theory suggests that individuals maximize their utility subject to a budget constraint. Accordingly, Sjaastad (1962) argues that the migration decision is based on the comparison between expected benefits and costs from migration. Potential migrants repeat this exercise for each potential destination country and choose the country that provides the best opportunities. The expected benefits and costs from migration depend on many factors related to the characteristics of the individual, the individual's origin country and those of all potential destination countries. In line with **?**, Pedersen et al. (2008) and Mayda (2010), we write aggregate migration from origin country o to destination country d at time t as a function of destination, origin and destination-origin characteristics capturing the benefits and costs of migration. Specifically, we define the aggregate migration rate as

$$\frac{M_{dot}}{N_{dt}} = \alpha_0 + \alpha_1 B_{dt} - \alpha_2 B_{ot} - \alpha_3 C_{dot} + \varepsilon_{dot} \tag{1}$$

$$B_{dt} = \ln(Y_{dt} Z_{dt}) \tag{2}$$

$$B_{ot} = \ln(Y_{ot} Z_{ot}) \tag{3}$$

where the migrant flow, M_{dot}, is divided by the resident population in the destination country, N_{dt}, to account for scale effects related to the fact that larger countries are able to provide more opportunities to and host more immigrants.[3] B_{dt}, B_{ot} and C_{dot} denote the expected benefits from migrating to destination d, those for staying in the home country o and the expected costs from migration from o to d, respectively. The expected benefits from migration or staying in the home country are a function of average incomes, Y, and the non-monetary returns, Z, while ε_{dot} denotes the error term, which is assumed i.i.d.[4]

Following Todaro (1969) and Harris and Todaro (1970), expected income is defined as the average income (inc) times the employment rate ($empl$) to account for the risk of not finding a job upon arrival in the destination country. Yet, in line with Hatton (1995), we assume that expected earnings abroad are subject to more uncertainty than those in the home country.[5] In fact, we do not impose equal coefficients

[3]The empirical specification described in equation (1) can be formalised as a linear probability model, i.e. a linear approximation to a model describing the probability that an individual i from country o decides to migrate to d at time t. The corresponding linear probability model would be given by

$$\frac{M_{dot}}{N_{dt}} = Prob(m_{idot} = 1) = \alpha_0 + \alpha_1 B_{dt} - \alpha_2 B_{ot} - \alpha_3 C_{dot} + \varepsilon_{dot}.$$

This relationship allows the model to be fitted using simple linear regression techniques. As argued by Caudill (1988) and Angrist and Pischke (2008), a carefully chosen linear model can yield good estimates of marginal effects, despite some of the well-known drawbacks of the linear probability model. Whereas probit or logit models are generally preferred to a linear probability model, the former only prove better estimators when the disturbances are known to be normally or logistically distributed, respectively. Moreover, contrary to probit or logit models, a linear probability model permits estimation of country specific effects and the parameters are directly interpretable (see e.g. Verbeek, 2012).

[4]Section 4 demonstrates how we account for potential spatial dependence in the migratory process and how this affects the structure of the error term. As a robustness check, we also control for the presence of unobserved heterogeneity among destination and origin countries. The results are discussed in Section 5. There is not much sense in adding a time effect given that our sample is limited to two time periods.

[5]In fact, Hatton (1995) explicitly takes into account uncertainty about employment prospects abroad and expects a

for employment prospects and average incomes at home or abroad. Taking logarithms, we can write expected incomes in the destination and origin countries, respectively, as

$$\ln Y_{dt} = \beta_1 \ln inc_{dt} + \beta_2 \ln empl_{dt} \tag{4}$$

$$\ln Y_{ot} = \delta_1 \ln inc_{ot} + \delta_2 \ln empl_{ot}. \tag{5}$$

Combining equations (1), (4) and (5) gives

$$\begin{aligned} \frac{M_{dot}}{N_{dt}} &= \alpha_0 + \alpha_1\beta_1 \ln inc_{dt} - \alpha_2\delta_1 \ln inc_{ot} \\ &+ \alpha_1\beta_2 \ln empl_{dt} - \alpha_2\delta_2 \ln empl_{ot} \\ &+ \alpha_1 \ln Z_{dt} - \alpha_2 \ln Z_{ot} - \alpha_3 C_{dot} + \varepsilon_{dot}. \end{aligned} \tag{6}$$

Through the identification of Z_{dt}, Z_{ot} and C_{dot}, this basic human capital model of migration can be elaborated to account for more structural influences of migration.

First, a popular proxy for the cost of migration, C_{dot}, is the social network: family and friends already in the host country may lower the psychological cost for newcomers leaving their familiar surroundings, alleviate financial constraints or help finding a job or housing. To capture these network effects, we incorporate the lagged stock of immigrants already present in the host country (MST) (see also Hatton, 1995; Fertig, 2001; Lewer and Van den Berg, 2008; Pedersen et al., 2008). Also the distance between origin and destination country (*dist*) and the presence of a common border (*commbord*) are considered suitable proxies for monetary expenses and non-monetary opportunity costs (such as foregone earnings while traveling and finding a job) incurred by the migrant (Karemera et al., 2000; Gallardo-Sejas et al., 2006; Lewer and Van den Berg, 2008; Pedersen et al., 2008; Mayda, 2010). Other factors expected to lower the costs of migration are the presence of a common language (*commlang*) and a common colonial past (*commcol*). Cultural similarities in the host and source country are assumed to make adaptation to the new environment easier, which in turn increases the propensity to migrate between these countries (see also Karemera et al., 2000; Gallardo-Sejas et al., 2006; Lewer and Van den Berg, 2008; Pedersen et al., 2008). Furthermore, we also investigate the impact of regional integration (*regint*) on migration. A positive sign might indicate that regional integration succeeds in stimulating the free movement of people whereas a negative sign might point to a substitution relationship between trade and labor. As such, the cost of migration, C_{dot}, is specified as

$$\begin{aligned} C_{dot} &= \varphi_0 + \varphi_1 \ln MST_{dot-1} - \varphi_2 \ln dist_{do} \\ &+ \varphi_3 commbord_{do} + \varphi_4 commlang_{do} + \varphi_5 commcol_{do} + \varphi_6 regint_{do} \end{aligned} \tag{7}$$

Second, we look more closely into specific characteristics of the origin and destination countries, Z_{dt} and Z_{ot}, which are likely to influence the return to migration and as such also the decision to migrate. Following the standard practice in the literature, the immigrant's income perspectives in the host country

higher coefficient for employment in the destination compared to the origin country. The same reasoning could be applied to the coefficients for other variables such as wages or education prospects. Whereas Hatton (1995) assumes that the probability of employment follows a biniomial distribution, we do not assume any specific distribution and do not impose any restrictions on the coefficients.

are proxied by GDP per capita. Borjas (1989) and Mayda (2010), however, argue that this proxy does not signal the true income opportunities for an immigrant because differences between the GDP per capita in host and source country are affected by differences in skill intensity. To capture this and to control for the effect of skill differences on GDP per capita, we follow Borjas (1989) and Mayda (2010) by adding the mean skill level of the population (*educ*) in the destination and origin country to the model. We expect the first (latter) to have a negative (positive) impact on migration. Next, assume the decision to migrate does not only depend on the current utility difference net of migration costs, but also on the net present value of all future ones. Specifically, the expected returns to migration are discounted over the remaining lifetime and therefore decreasing with age. As such, young people will have more incentive to migrate, because the discounted value of their expected returns is higher due to their longer remaining working life. Following the literature, we control for this effect by incorporating the share of the young population in the origin country (*youngpop*) (see also Hatton and Williamson, 2002; Gallardo-Sejas et al., 2006; Mayda, 2010). Symmetrically, we include the share of the young population in the destination to capture tension in the host country's labor market, which provides an indication of job opportunities for migrants. Migration is expected to be higher the larger (smaller) the share of young people in the origin (destination) country. Finally, we account for the non-monetary return of migration that arises from locational characteristics, such as sociopolitical and environmental circumstances. Obviously, migrants are expected to prefer countries with less conflict ($confl$) and more relative freedom (fr) (Karemera et al., 2000; Hatton and Williamson, 2002; Pedersen et al., 2008). The latter combines measures of civil liberties and political rights. Because the freedom status and its components are all highly correlated, we include only the former. Additionally, extreme conditions caused by (natural) disasters (*disaster*) or weather anomalies (*climate*) have proven to affect especially the poorest and powerless, for whom migration might be one of many coping mechanisms (Barrios et al., 2006). It is expected that people are more (less) likely to move away from (towards) countries affected by disaster and extreme temperature (see Findley, 1994; Ezra and Kiros, 2001). Hence, $\ln Z_{dt}$ and $\ln Z_{ot}$ are specified as

$$\begin{aligned} \ln Z_{dt} &= \gamma_0 + \gamma_1 \ln educ_{dt} - \gamma_2 \ln popyoung_{dt} - \gamma_3 \ln confl_{dt} \\ &\quad + \gamma_4 \ln fr_{dt} - \gamma_5 \ln disaster_{dt} - \gamma_6 \ln climate_{dt} \qquad (8) \\ \ln Z_{ot} &= \eta_0 + \eta_1 \ln educ_{ot} + \eta_2 \ln popyoung_{ot} + \eta_3 \ln confl_{ot} \\ &\quad - \eta_4 \ln fr_{ot} + \eta_5 \ln disaster_{ot} + \eta_6 \ln climate_{ot} \qquad (9) \end{aligned}$$

Replacing in (6) C_{dot}, Z_{dt} and Z_{ot} by their components and regrouping yields a comprehensive empirical specification of the human capital migration model given by

$$\begin{aligned}
\frac{M_{dot}}{N_{dt}} &= (\alpha_0 + \varphi_0 + \alpha_1\gamma_0 + \alpha_2\eta_0) \\
&\quad + \alpha_1\beta_1 \ln inc_{dt} - \alpha_2\delta_1 \ln inc_{ot} + \alpha_1\beta_2 \ln empl_{dt} - \alpha_2\delta_2 \ln empl_{ot} \\
&\quad + \alpha_3\varphi_1 \ln MST_{dot-1} - \alpha_3\varphi_2 \ln dist_{do} + \alpha_3\varphi_3 commbord_{do} \\
&\quad + \alpha_3\varphi_4 commlang_{do} + \alpha_3\varphi_5 commcol_{do} + \alpha_3\varphi_6 regint_{do} \\
&\quad - \alpha_1\gamma_1 \ln educ_{dt} + \alpha_2\eta_1 \ln educ_{ot} \\
&\quad - \alpha_1\gamma_2 \ln popyoung_{dt} + \alpha_2\eta_2 \ln popyoung_{ot} \\
&\quad - \alpha_1\gamma_3 \ln confl_{dt} + \alpha_2\eta_3 \ln confl_{ot} + \alpha_1\gamma_4 \ln fr_{dt} - \alpha_2\eta_4 \ln fr_{ot} \\
&\quad - \alpha_1\gamma_5 \ln disaster_{dt} + \alpha_2\eta_5 \ln disaster_{ot} \\
&\quad - \alpha_1\gamma_6 \ln climate_{dt} + \alpha_2\eta_6 \ln climate_{ot} + \varepsilon_{dot}. \qquad (10)
\end{aligned}$$

Like in Hatton (1995), Pedersen et al. (2008) and Mayda (2010), our model of international migration has a semi-log functional form, which has the important advantage that it allows to explain not only positive, but also zero and even negative migration rates. This point will be relevant for the choice of our estimation method discussed in Section 4.

On the whole, the empirical specification accounts for the traditional *economic determinants*, reflecting average incomes through wages and employment opportunities; *network effects*, captured by the stock of immigrants from the same ethnic origin already in the host country; *geographical and cultural proximity*, measured by the distance between the origin and destination country, the presence of a common border, a common language, a common colonial past and a proxy for regional integration; *the demographic situation*, proxied by the level of education and the share of the young population in the total population; *the political situation* through the occurrence of conflict and citizens' relative freedom; and finally *the environmental impact* captured by the incidence of disaster and the severity of temperature anomalies.

3 The data

As argued in the introduction, the lack of complete and reliable data has formed a major obstacle for an in-depth analysis of the incentives for South-South migration. In many developing countries, and particularly in SSA, keeping track of migratory streams has not been a major concern. Organizations such as the United Nations and the US Census Bureau provide estimates on long-term international migration in SSA. Yet, these figures do not allow for a South-South analysis since they are not disaggregated by country of origin.

The approach of early studies of immigration between countries as well as studies of internal movements was to define their dependent variable as the number of persons, born in a given place of origin, residing in each of the destination localities at the date of the census. That is, a migrant stock, rather than a flow variable was used. As a result no distinction could be made between recent and earlier migrants or between those who settled directly in the observed destination and those who arrived through a succession of moves. Furthermore, the migrant stock reflects the result of a process taking place over many years, while the explanatory variables are usually measured at one point in time. Consequently, the determinants may not reflect the conditions existent at the time of the actual move (Dunlevy, 1980).

The recently constructed GBMD, on the other hand, offers the opportunity to create migration flows in three dimensions (destination, origin and time), which allow for a rigorous analysis of the determinants of South-South migration. It builds on the United Nations Population Division's Global Migration Database, which augments and updates the bilateral migration matrix compiled by the University of Sussex and Ratha and Shaw (2007). The database mostly provides statistics on foreign born wherever possible, and foreign nationals otherwise. Although the migrant stock data is not perfectly comparable across countries, substantial effort has been made to standardize the data and ensure consistent figures for the number of migrants in each of the five census periods. Though migration on the African continent is in part irregular (given ill-defined migration laws and inconsistent border control), it provides a fairly accurate picture of migratory movements during the period (Beine et al., 2011).[6]

Based on this database, we define our dependent variable as the change in bilateral migrant stocks, that is the difference in the number of foreign residents in each country disaggregated by country of origin, for each decade between 1980-2000. The change in migrant stocks is divided by the population in the destination country (in thousands) to control for size effects as described in Section 2.

Given that migration between SSA and northern Africa is very small (the World Bank reports not a single SSA migrant in North-Africa and also in the other direction there is little border crossing) and mainly consists of transit migration, we exclude the north African countries from our sample. Furthermore, also Djibouti, Eritrea, Mayotte, Saint Helena, Sao Tome and Principe, Reunion, the Seychelles, Sudan and Western Sahara are dropped because of missing information for certain country characteristics. For the same reason, our sample is limited to the last two decades in the database. Finally, our sample contains statistics on the change in the stock of migrants in 42 destination countries from the same 42 origin countries, between 1980-1990 and 1990-2000.[7]

Appendix 8.1 documents detailed information on measurement and data sources for the explanatory variables used in the empirical model.[8] The data have been compiled from various international organizations and research institutes like the World Bank, the United Nations and CEPII. As such, our dataset enables us to proxy for all the determinants used in the empirical model described in Section 2. As an indicator of the beginning-of-period values of the explanatory variables, we take the average over the 5-year period prior to the start of the corresponding decade unless stated otherwise. As such, we repress potential problems of endogeneity bias and erratic deviance from the trend value. In the same vein, we use lagged values of the migrant stock, that is the observation prior to the corresponding decade for the dependent variable (we cannot take 5-year averages for migrant stocks because the data are available only decennially).

[6] It is worth mentioning that refugees in camps have been excluded from the database to make the distinction between refugee flows and actual migration. For explicit details on how the data on migrant stocks have been collected and harmonized, we refer to Özden et al. (2011).

[7] For an overview of migration stocks and changes by destination and origin, see Tables 4 and 5.

[8] Summary statistics can be found in Table 6.

4 Spatial dependence and estimation method

A model of bilateral migration, like (10), can be considered a 'spatial interaction model', i.e. a model that focuses on flows between origins and destinations as described in Sen and Smith (1995). These models typically explain bilateral flows as a function of characteristics of both origin and destination regions as well as the distance between them. Also the gravity model belongs to this family with several applications in the migration literature (Karemera et al., 2000; Ortega and Peri, 2009). Yet, all of the existing models assume independence of observations, which might be problematic in several contexts, and the recognition of the need to account for spatial dependence in analyzing human migration is widespread (Cushing and Poot, 2004; LeSage and Pace, 2008, 2009; Mitze, 2009).

Using distance as an explanatory variable, gravity models do not effectively capture spatial dependence in international flows (Curry, 1972; Griffith, 2007; LeSage and Pace, 2009). In cases where each country might affect its neighbors, this approach proves inadequate because it ignores the spatial interrelatedness of bilateral flows. The spatial econometrics literature provides both theoretic and econometric motivations for the use of spatial regression models. An example of the former concerns migration regulations, which are difficult to measure in practice because of their qualitative nature and, therefore, often omitted in empirical specifications. They form, however, an important barrier to migration and are likely to be correlated across countries. Governments might, for instance, decide to set in place certain policy measures after having observed those set by neighboring countries. This type of spatial interdependence might be explicitly integrated in the formal specification of the theoretical model. Yet, it might also be motivated from an econometric perspective by looking upon bilateral flows as describing a diffusion process over space with a time lag. This form of spatial dependence typically shows up in cross-sectional models with a spatial lag of the dependent variable. Another important econometric motivation for the use of spatial regressions concerns the presence of omitted latent influences that are spatial in nature, typically leading to a spatial Durbin model (SDM) with spatial lags of both the dependent and explanatory variables (LeSage and Pace, 2009). Again, migration policy appears an obvious candidate given that it is often an omitted latent influence that is both correlated with the explanatory variables and across locations. Especially the second of these econometric motivations is relevant in the context of this chapter.[9]

LeSage and Pace (2009) show that the SDM is less affected by omitted variable bias than a model that ignores spatial dependence. This holds when the omitted variable is truly involved in the data generating process, but also when it is not, its inclusion does not lead to bias in the estimates. Consequently, the authors suggest relying on a model that includes spatial lags of the dependent and explanatory variables even if this seems counter to the underlying theory behind our model. Note that we do not a priori impose any spatial dependence in the migrant flow, as this does not immediately follow from current theoretical models motivated by utility considerations. In line with LeSage and Pace (2008, 2009), our starting point is consistent with the human capital model, which posits a non-spatial theoretical relationship underlying migration flows.

[9]The first econometric motive is less likely in view of the time span (10 years) over which we consider the migration rates.

In a model of bilateral flows (like international trade or migration), the spatial interaction structure is likely to be more complex compared to standard spatial lag or spatial error models, because it needs to take into account spatial correlation of the flows at both origins and destinations (LeSage and Pace, 2008, 2009). To emphasize the origin-destination (OD) structure of the migration model, rewrite the unrestricted form of equation (10) as

$$\frac{M_{dot}}{N_{dt}} = \theta_0 + \theta_1 \ln MST_{dot-1} + X_{dt}\theta_2 + X_{ot}\theta_3 + X_{do}\theta_4 + \varepsilon_{dot} \tag{11}$$

where X_{dt} denotes time-varying destination characteristics, X_{ot} time-varying origin characteristics, X_{do} time invariant bilateral characteristics, $\theta_0 = \alpha_0 + \varphi_0 + \alpha_1\gamma_0 + \alpha_2\eta_0$, $\theta_1 = \alpha_3\varphi_1$, $\theta_2 = \alpha_1 (\beta_1\beta_2\gamma_1...\gamma_6)'$, $\theta_3 = \alpha_2 (\delta_1\delta_2\eta_1...\eta_6)'$, $\theta_4 = \alpha_3(\varphi_2...\varphi_6)'$. Subsequently, we add spatial lags for both the dependent and explanatory variables using a combination of three spatial connectivity matrices W_d, W_o and W_w, for destination, origin and destination-to-origin dependence respectively, as suggested by LeSage and Pace (2008, 2009). The spatial weight matrices are row-normalized contiguity matrices of order one, which take a positive value when two countries are neighbors and zero otherwise. This results in the unconstrained SDM model,

$$\begin{aligned}\frac{M_{dot}}{N_{dt}} = {} & \theta_0 + \rho_d W_d \left(\frac{M_{dot}}{N_{dt}}\right) + \rho_o W_o \left(\frac{M_{dot}}{N_{dt}}\right) + \rho_w W_w \left(\frac{M_{dot}}{N_{dt}}\right) \\ & + \theta_1 \ln MST_{dot-1} + X_{dt}\theta_2 + X_{ot}\theta_3 + X_{do}\theta_4 \\ & + \theta_5 W_w \ln MST_{dot-1} + W_d X_{dt}\theta_6 + W_o X_{ot}\theta_7 + W_w X_{do}\theta_8 + \varepsilon_{dot}\end{aligned} \tag{12}$$

the most general form of spatial dependence. Subsequently, we run a series of Wald tests to determine whether the SDM can be simplified to a spatial lag or a spatial error model.

LeSage and Pace (2008, 2009) propose a maximum likelihood estimator (MLE) to estimate the SDM. In the context of bilateral migration flows, however, it might be argued that a large number of zero flows invalidates the normality assumption needed for maximum likelihood estimation. For our sample of 1880-2000 migration rates, we have zero values in about 33 percent of the observations. One suggestion to address the issue of zero flows is to aggregate the data to larger spatial units or cumulating flows over a longer time period. Our current database however already considers flows at the highest level of aggregation, that is the country level, which are obtained by combining flows over 10 year periods. Moreover, the fact that our dependent variable also takes negative values (for instance in cases where return migration exceeds immigration between two countries) prevents us from using count data methods such as multinomial logit or tobit models, which by definition require non-negative values (see Beine et al., 2011). The semi-log functional form of our empirical model however allows us to explain migration flows, irrespective of their sign.

To account for the non-normality of the migrant rate, we estimate the empirical SDM using a quasi-maximum likelihood estimator (QMLE), which produces consistent estimates, even if the likelihood function is not entirely correct (but the first-order conditions are) (see White, 1982; Verbeek, 2012). The information matrix test developed by White (1982), suggests that the distribution of the QMLE differs from that of the MLE. The small sample distribution of the QMLE can however be obtained in a numerical way by resampling the original data a 1000 times and applying the MLE in each of the constructed samples. By resampling the data within but not between cross-section units, the data resampling procedure aligns with the assumed data generating process of the data. As such, inference is based on the simulated distribution of the QMLE which allows us to calculate robust standard errors and t-statistics.

An alternative methodology suggested by LeSage and Pace (2009) concerns a Markov Chain Monte Carlo (MCMC) approach, which is based on the decomposition of the posterior distribution into a set of conditional distributions for each parameter in the model. Bayesian parameter estimates are then obtained from repeated sample draws from these conditionals. This approach has the advantage that it decomposes a complicated estimation problem into simpler problems without having to carry out numerical integration of the posterior distribution with respect to the parameters as was needed in conventional Baysian methodology. It is however still considered quite controversial given the subjective choice of prior distributions, the lack of an objective principle for choosing a non-informative prior and the potential influence of these choices on the estimation outcome. Moreover, MCMC techniques cannot guarantee that convergence has taken place. To check the robustness of our results, we re-estimated our empirical model using the MCMC approach suggested by LeSage and Pace (2009) and obtained similar results compared to the QML estimates discussed below.[10]

An implication of accounting for spatial dependence is that the estimated parameters cannot be interpreted as usual in a standard linear regression model. Cross-country interactions prevent the parameter estimates from being interpreted as the simple partial derivatives of the dependent variable with respect to the explanatory variables (see Anselin and Le Gallo, 2006; Kelejian et al., 2006; LeSage and Pace, 2009). Pace and LeSage (2006) and LeSage and Pace (2009) suggest three summary measures of the varying impacts of changes in an explanatory variable across countries:

(i) average direct impact: the impact from changes in the ith observation of variable k on country i, averaged over all countries

(ii) average indirect impact: the effect of changes in the ith observation of variable k on country j ($\neq i$), averaged over all countries, capturing the spillover effects of a change in country i on all other countries

(iii) average total impact: the sum of the previous two, reflecting how changes in a single country potentially influence all observations.

The direct effects correspond the most to the typical regression coefficient interpretation that represents the average response of the dependent variable to independent variables over the sample of observations. The main difference is that the direct effect takes into account feedback effects from changes in country i to country j and back to country i itself. Because they allow for an explicit comparison with parameter estimates from other studies on migration determinants in the literature, we will concentrate primarily on the average direct effects in the discussion of our results, although we will also consider the indirect

[10] The MCMC estimation results for the SDM model are available upon request from the authors. Although QMLE puts more (less) emphasis on the spatial lags of the dependent variable (explanatory variables) compared to MCMC, the estimated direct and total effects are fairly similar across estimation methods.

effects[11] and briefly comment upon the total effects.

The various types of effects estimates are calculated using the empirical distribution of the model parameters. The latter is constructed using a large number of simulated parameters drawn from the QML multivariate normal distribution of the parameters as suggested by LeSage and Pace (2009). Using a 1000 simulated draws, we compute means, standard deviation and t-statistics for direct, indirect and total impacts. For technical details on the calculation of these summary measures as well as measures of dispersion for the impact estimates, we refer to LeSage and Pace (2009).

5 Estimation results

In what follows, we present estimation results for nine models in which specific categories of variables are added sequentially until the complete model is reached in the final column.[12] As argued above, we perform a number Wald tests to decide whether the SDM model can be simplified to a spatial lag or spatial error model. The latter are rejected in favor of the SDM, suggesting that the most appropriate model is the one that includes spatial lags of both the dependent and the explanatory variables. Table 1 displays test statistics and p-values for each of the nine models. Starting from the basic human capital model with economic determinants and network effects and sequentially adding geographical, cultural, demographical, sociopolitical and environmental explanatory variables, we are able to explain nearly 60 percent of the variation in migration streams.[13]

Based on these test results, all of the nine SDM models are estimated using pooled QML with three sources of spatial dependence. From Table 1, we see that both the destination-based and origin-based spatial lags of the dependent variable are statistically significant, with a dominant influence from the latter. The destination-to-origin based spatial lag, on the other hand, appears insignificant. This suggests the presence of both destination and origin spatial dependence in the migration flow between SSA countries during 1980-2000.[14]

Tables 3 and **??** report the summary measures of the SDM direct and indirect effects for each of the nine models.[15] With a few exceptions, our results are fairly robust across specifications and mostly consistent with the theoretical predictions of the international migration model.

[11]Technically, for the kth variable, the average direct (indirect) effect corresponds to the average of the main diagonal (the average of the row sums of the off-diagonal) elements of the matrix $(I - \rho W)^{-1} \left(I\theta_{i,k} + W\theta_{i,k+4}\right)$ in (12).

[12]To be able to estimate a panel version of the SDM using three connectivity matrices, we combined the Matlab software for spatial panels provided by Elhorst (2010, 2013) at his website and the spatial econometric modelling of origin-destination flows described in LeSage and Pace (2008, 2009).

[13]The log likelihood function is likely to be misspecified due to the non-normality of the residuals. Therefore, we cannot rely on likelihood ratio tests to determine whether our general model could be simplified to one of the nine more specific models set forth in LeSage and Pace (2008) which impose various restrictions on the parameters for the spatially lagged dependent variable. Yet, considering that the inclusion of insignificant spatial lags will not lead to bias (see above), we prefer to use the most general model 9 in LeSage and Pace (2008) in all of our model specifications.

[14]The remaining parameter estimates together with their simulated t-statistics for these models can be found in Table 8. The difference between the parameter estimates and the direct effects estimates is due to feedback effects that arise as a result of impacts passing through neighboring countries and back to the country itself (see LeSage and Pace, 2009).

[15]Given that the estimated total effects are simply the sum of estimated direct and indirect effects, the latter are not reported in the text, but can be found in Table 9.

Table 1: Spatial Durbin model estimates

Dependent variable: $\ln M_{dot}/\ln N_{dt}$								Sample period: 1980-2000	
	I	II	III	IV	V	VI	VII	VIII	IX
Log likelihood	-10492	-7824	-7519	-7382	-7392	-7338	-8921	-8520	-7125
Corr^2	0.050	0.011	0.077	0.252	0.510	0.525	0.527	0.584	0.664
Adjusted R^2	-0.174	-0.145	-0.011	0.190	0.422	0.434	0.343	0.423	0.583
Wald Spatial Lag	0.734	8.414	8.622	5.294	9.396	8.945	9.089	11.610	19.230
$Prob > \chi^2$	0.693	0.015	0.071	0.381	0.402	0.537	0.825	0.867	0.631
Wald Spatial Error	3.065	6.998	4.225	3.099	7.801	7.545	8.125	12.007	19.415
$Prob > \chi^2$	0.216	0.030	0.376	0.685	0.554	0.673	0.883	0.847	0.620
$W_d M_{dot}$	0.012^{***}	0.016^{***}	0.017^{***}	0.017^{***}	0.017^{***}	0.017^{***}	0.017^{***}	0.007^{**}	0.018^{***}
	(3.259)	(4.106)	(4.734)	(4.498)	(4.678)	(4.497)	(4.398)	(1.983)	(4.587)
$W_o M_{dot}$	0.285^{***}	0.239^{***}	0.185^{***}	0.182^{***}	0.182^{***}	0.181^{***}	0.263^{***}	0.233^{***}	0.158^{***}
	(7.387)	(6.473)	(4.893)	(4.746)	(4.827)	(4.743)	(7.161)	(6.242)	(4.348)
$W_w M_{dot}$	-0.017	0.009	0.025	0.026	0.025	0.026	0.010	-0.035	0.032
	(-0.747)	(0.371)	(1.098)	(1.109)	(1.082)	(1.109)	(0.435)	(1.575)	(1.391)

Notes: T-statistics in parenthesis are calculated using simulated standard errors based on 1000 bootstrap samples. *** and ** indicate significance at the 1% and 5% level, respectively. Number of observations: 3444. 'Corr^2' denotes the correlation coefficient between actual and fitted values. 'Adjusted R^2' is the coefficient of determination corrected for the degrees of freedom.

Focussing on the direct effects first, we start by regressing migrant rates on the economic determinants. Regressions I to IV suggest positive (negative) significant direct effects for income in the destination (origin) country, in line with our expectations, but ambiguous effects for employment rates. In fact, the positive significant impact of income in the destination country is the most robust result across specifications. Income in the origin country has a significantly negative direct effect on migration rates in regression I but this effect diminishes in model III when employment rates enter the equation. When introduced separately, employment rates have an insignificant direct effect on migration rates. Yet, the estimated impact of employment in the origin country becomes significantly negative once we control for average income, in line with the predictions of the human capital model.

Regression IV introduces the network effect. As expected, we find a positive and highly significant effect of migrant stocks on migration rates. According to the estimates in regression IV, an increase in the lagged bilateral migrant stock by 100 persons on average attracts another 7 persons per 1000 individuals in the destination from the same origin. Though these effects are rather small, they provide some first evidence for the role of network effects in encouraging migratory streams in SSA. Ignoring economic determinants or introducing them separately together with the lagged stock variable does not

Table 2: Direct effects estimates

Dependent variable: $\ln M_{dot} / \ln N_{dt}$								Sample period: 1980-2000
Variable	I	II	III	IV	V	VI	VII	VIII	IX
$\ln inc_{dt}$	0.102^{***}		0.163^{***}	0.154^{***}	0.201^{***}	0.212^{***}	0.240^{***}	0.253^{***}	0.242^{***}
	(2.893)		(3.109)	(3.064)	(3.475)	(3.677)	(3.278)	(3.359)	(3.300)
$\ln inc_{ot}$	-0.095^{***}		-0.070^{**}	-0.043	-0.011	0.000	0.004	0.001	-0.012
	(-3.015)		(-2.154)	(-1.338)	(-0.296)	(-0.007)	(0.087)	(0.024)	(-0.277)
$\ln empl_{dt}$		-0.041	-0.004	-0.051	0.071	0.070	0.037	0.021	0.030
		(-0.463)	(-0.033)	(-0.468)	(0.622)	(0.551)	(0.354)	(0.222)	(0.244)
$\ln empl_{ot}$		0.011	-0.192^{***}	-0.214^{**}	-0.141	-0.132	-0.144	-0.160^{*}	-0.095
		(0.116)	(-2.072)	(-2.173)	(-1.340)	(-1.270)	(-1.591)	(-1.735)	(-0.758)
$\ln MST_{dot-1}$				0.069^{***}	0.018	0.023^{*}	0.024	0.027^{*}	0.032^{**}
				(3.367)	(1.575)	(1.776)	(1.636)	(1.799)	(2.241)
$\ln educs_{dt}$							-0.068	-0.082^{*}	-0.131^{**}
							(-1.449)	(-1.750)	(-2.024)
$\ln educs_{ot}$							-0.011	-0.018	-0.022
							(-0.253)	(-0.282)	(-0.355)
$\ln youngpop_{dt}$							-0.050	-0.083	0.062
							(-0.163)	(-0.284)	(0.238)
$\ln youngpop_{ot}$							0.119	0.215	-0.001
							(0.294)	(0.489)	(-0.004)
$\ln confl_{dt}$								0.020	0.243
								(0.148)	(1.528)
$\ln confl_{ot}$								0.205	0.212
								(1.271)	(1.374)
$\ln fr_{dt}$								0.000	0.078
								(0.000)	(0.400)
$\ln fr_{ot}$								0.102	0.127
								(0.695)	(0.883)
$\ln disaster_{dt}$									-0.033^{*}
									(-1.931)
$\ln disaster_{ot}$									-0.012
									(-1.639)
$\ln climate_{dt}$									0.114
									(0.288)
$\ln climate_{ot}$									-0.117
									(-0.787)
$\ln distance_{do}$					-0.154^{*}	-0.174^{*}	-0.183	-0.204	-0.193
					(-1.672)	(-1.848)	(-1.467)	(-1.499)	(-1.400)
$commbord_{do}$					0.607^{**}	0.602^{**}	0.593^{**}	0.573^{**}	0.551^{**}
					(2.271)	(2.266)	(2.326)	(2.209)	(1.995)
$commcol_{do}$					-0.001	0.014	0.015	0.017	0.034
					(-0.009)	(0.137)	(0.136)	(0.139)	(0.296)
$commlang_{do}$					0.076	0.051	0.073	0.058	0.038
					(0.879)	(0.579)	(0.737)	(0.692)	(0.518)
$regint_{do}$						-0.150	-0.153	-0.183^{*}	-0.184
						(-1.452)	(-1.414)	(-1.667)	(-1.490)

Notes: T-statistics in parenthesis are calculated using simulated standard errors based on 1000 bootstrap samples. *** and ** indicate significance at the 1% and 5% level, respectively. Number of observations: 3444.

alter this finding (not reported here). Given that our model includes destination, origin and destination-to-origin dependence, this implies that an increase in migrant stocks between one pair of destination and origin countries not only affects migration rates in the respective destination but also in neighbors to this destination and in neighbors to the countries where the migration flows originate. The same reasoning can be applied to spillover effects arising from changes in the other explanatory variables.

In the subsequent regressions, we explore the role played by geographical and cultural proximity. It becomes immediately clear that both distance and especially the presence of a common border are fairly important and robust factors for explaining migration rates across specifications (see also Karemera et al., 2000; Mayda, 2010, for the effect of distance). In line with our expectations, migration rates decrease with distance (significantly across all models when we would apply a one-sided test) and are higher when two countries share a common land border. The impact of past colonial relationships appears statistically insignificant. The same holds for the presence of a common language suggesting that, when we control for the other variables included in the regression, cultural proximity does not appear to affect migration rates (see also Mayda, 2010). It should be noted that controlling for geographical and cultural proximity slightly alters the picture. First, it removes the statistically significant direct effect of employment in the origin country from regression IV. Second, it reduces the estimated parameter and significance of the network effect. To be more precise, the direct coefficients for the migrant stock show a substantial drop when we control for bilateral effects but then gradually recover once also demographics, sociopolitical characteristics and especially environmental factors are taken into account.

Regression VI introduces regional integration in the estimation equation. Although we find unambiguous negative direct effects across specifications, the estimated impact is only marginally significant in model VII. As such, we do not find evidence for a positive influence of regional integration on migration through the enhancement of free movement, nor for a negative influence linked to substitution between trade and labor as discussed above.

Next, we introduce the demographic variables. We find that the migration rate is negatively related to the level of secondary education in the destination. This supports the argument of Borjas (1989) and Mayda (2010) for the necessity to correct for the effect of skill differences on the proxies for the immigrant's income perspectives, at least in the destination country. The schooling level in the origin country, on the other hand remains insignificant (in line with Mayda, 2010). As far as concerns the share of the young population, though generally of the right sign, we find insignificant effects. As such, we cannot confirm that intraregional migration in Sub-Saharan Africa responds to fluctuations in employment due to demographic pressure, or that the incentive to migrate significantly decreases with age.

In regressions VIII and IX we investigate to what extent migration rates are shaped by the sociopolitical characteristics of origin and destination countries. We find no evidence of an important role played by these factors (except for the occurrence of conflict in the source country in a one-sided test).

Finally, regression IX combines all regressors described above and investigates the relative importance of environmental factors in explaining the migration rate. According to our estimations, the number of people affected by disaster relative to the population in the host country has a significantly negative direct effect on migration. Hence, the destination choice of immigrants is influenced by the occurrence of (natural) disaster. Our evidence suggests insignificant coefficients for the remaining direct effects after controlling for other aspects of the migration decision. Robustness checks using more specific proxies for the environmental impact, such as the relative number of people affected by natural disasters (drought, earthquake, epidemic, extreme temperature, flood, insect infestation, mass movement, storm, volcano or wildfire) or climatic disasters in particular (drought, extreme temperature or wildfire), and even using the

Table 3: Indirect effects estimates

Dependent variable: $\ln M_{dot} / \ln N_{dt}$ — Sample period: 1980-2000

Variable	I	II	III	IV	V	VI	VII	VIII	IX
$\ln inc_{dt}$	0.002		−0.043	−0.051	0.020	0.034	0.080	0.070	0.001
	(0.293)		(−1.361)	(−1.387)	(0.282)	(0.484)	(0.777)	(0.596)	(0.009)
$\ln inc_{ot}$	−0.028		−0.041	−0.022	0.015	0.028	0.039	0.000	−0.040
	(−1.586)		(−0.643)	(−0.398)	(0.340)	(0.591)	(0.300)	(0.002)	(−0.297)
$\ln empl_{dt}$		0.022**	0.109*	0.035	0.342*	0.322	0.069	0.333	0.674**
		(2.216)	(1.681)	(0.505)	(1.671)	(1.570)	(0.307)	(1.500)	(2.091)
$\ln empl_{ot}$		0.052*	0.042	0.013	−0.029	−0.046	0.214	0.139	0.019
		(1.644)	(0.367)	(0.121)	(−0.319)	(−0.473)	(1.094)	(0.750)	(0.108)
$\ln MST_{dot-1}$				0.083	0.010	0.010	0.013	0.033	0.065
				(1.201)	(0.260)	(0.248)	(0.277)	(0.577)	(0.985)
$\ln educs_{dt}$							−0.104	−0.154	−0.253
							(−0.884)	(−0.897)	(−1.292)
$\ln educs_{ot}$							0.067	0.106	0.065
							(0.479)	(0.690)	(0.491)
$\ln youngpop_{dt}$							0.286	0.038	−0.040
							(0.607)	(0.093)	(−0.085)
$\ln youngpop_{ot}$							−0.327	−0.202	−0.198
							(−1.243)	(−0.751)	(−0.926)
$\ln confl_{dt}$								−0.301	0.037
								(−0.910)	(0.117)
$\ln confl_{ot}$								0.632*	0.626*
								(1.698)	(1.733)
$\ln fr_{dt}$								0.562	0.792**
								(1.550)	(2.163)
$\ln fr_{ot}$								0.050	0.193
								(0.331)	(1.111)
$\ln disaster_{dt}$									−0.071***
									(−2.829)
$\ln disaster_{ot}$									−0.025
									(−1.196)
$\ln climate_{dt}$									−0.467
									(−0.947)
$\ln climate_{ot}$									0.505
									(1.401)
$\ln distance_{do}$					−0.202	−0.201	−0.205	−0.188	−0.212
					(−1.397)	(−1.357)	(−1.218)	(−1.204)	(−1.312)
$commbord_{do}$					1.249	1.214	1.151	0.919	0.946
					(1.565)	(1.568)	(1.475)	(1.373)	(1.370)
$commcol_{do}$					0.073	0.051	−0.016	−0.064	−0.051
					(0.214)	(0.149)	(−0.053)	(−0.249)	(−0.186)
$commlang_{do}$					−0.112	−0.069	−0.064	−0.121	−0.228
					(−0.367)	(−0.231)	(−0.204)	(−0.385)	(−0.628)
$regint_{do}$						−0.159	−0.155	−0.241	−0.300
						(−0.762)	(−0.684)	(−0.968)	(−1.199)

Notes: T-statistics in parenthesis are calculated using simulated standard errors based on 1000 bootstrap samples. *** and ** indicate significance at the 1% and 5% level, respectively. Number of observations: 3444.

number of people affected by these type of disasters in absolute terms do not alter these results. It should however be mentioned that the impact of natural disasters depends on the socio-economic situation of the people affected and, more specifically, on their adaptation mechanisms which improve their ability to cope with extreme climatic events (Meze-Hausken, 2000; Haug, 2002). Our variable capturing climate change, measured by temperature anomalies, appears insignificant. These results are robust to whether disaster and temperature anomalies enter the regression together or one at a time (not reported).[16]

As far as concerns the indirect effects, the estimates suggest that only a limited number of spillover effects are significant. First of all, the indirect effects of income per capita in the home and destination country are fairly small and always insignificant. In some models, we find significant positive spillover effects for employment rates in the destination country (in the model including only employment rates and the complete model IX), but this result is not robust. Hence, the economic determinants of migration (and migrant networks) only have a direct impact on international migration. For conflict at the origin country, however, we find a positive significant indirect impact, pointing to a regional dimension of conflict: the occurrence of wars in neighboring countries seems to incite people to leave their home country. In the last model, we also find evidence for a positive impact of relative freedom (in terms of civil liberties and political rights) in the broader destination area (in line with Barkley and McMillan, 1994), just as the occurrence of disaster in this area indirectly discourages migration towards the countries in that region. Therefore, apart for the economic determinants, we find evidence for the presence of spillover effects (and hence a regional dimension) only for the sociopolitical and environmental determinants in our model.

As mentioned above, the total effects are calculated as the sum of the direct and indirect effects. Given that for most variables, the latter remain fairly limited, the total effects estimates are generally very similar to those obtained for the direct effects (for instance for income per capita in the destination or origin country). An exception concerns the sociopolitical and environmental variables, for which the indirect effects significantly reinforce the direct effects, such that the total effects are substantially stronger than the latter.

Robustness checks

Next, we verify the robustness of our results for potential measurement error and unobserved heterogeneity. First, in case of measurement error, our results would be biased downward. In order to get an idea of potential measurement errors, we can exploit the time dimension of our data. Assuming that the problem of measurement error is the most serious for the oldest data (considering the efforts by international institutions in collecting data on developing countries in the recent decades), we re-estimated our model for each period separately (that is for migration flows between 1980-1990 and between 1990-2000, respectively), as a first robustness check. For the first period, we find relatively more coefficient estimates insignificantly different from zero compared to the panel data estimations. This is in line with what we would expect from measurement error. However, for the second period, we find significant coefficients of the same sign and results that are very similar to those obtained for the panel model. Because the

[16] We note that other studies have also used precipitation in their analysis of the impact of weather anomalies on migration rates. Adding rain anomalies would however imply a reduction in the sample size, which made us decide not to use it in our empirical analysis. Rainfall and temperature both drive evapotranspiration, suggesting that they might be considered alternative measures of the same event. Though different samples place different emphasis on the relative importance of rainfall or temperature, they find robust evidence for an impact from weather anomalies on migration (see e.g. Barrios et al., 2006; Marchiori et al., 2012). Others argue that crop growth and thus the impact of weather on agriculture income variability stems solely from temperature anomalies and should be measured accordingly (see e.g. Burke et al., 2009; Dillon et al., 2011).

estimations using only the more accurate data confirm the results of the overall estimation, we believe that the influence of measurement error in the reported results remains fairly mild.

Second, we re-estimate our model with destination and origin specific effects to test for the presence of unobserved heterogeneity. A number of Wald tests indicate that, for the most complete model, the hypotheses of jointly significant country specific effects can be rejected at the 1 percent significance level.[17] This suggests that there is no remaining unobserved heterogeneity once all categories of migration determinants as well as spatial interaction have been taken into account.

6 Conclusions

Despite great accomplishments in the migration literature, little is still known about the determinants of South-South migration. In an attempt to fill this gap, we examine what has been driving intraregional migration in SSA, using the World Bank's Global Bilateral Migration Database. We estimate the determinants of migration rates between 42 origin and destination countries for the period 1980-2000, taking into account spatial dependence in the migration decision.

Our theoretical framework is based on Sjaastad's (1962) human capital model of migration and encompasses economic variables as well as network effects, geographical and cultural proximity, demographics, the sociopolitical landscape and the environmental impact. This comprehensive model allows us to evaluate the relative importance of the different factors driving migration patterns in SSA. In addition, we allow for spatial dependence in the migration rates and their determinants. We find a significant impact of both destination- and origin-based spatial dependence in the migration decision, which confirms the necessity to control for both types of spatial correlation when estimating a bilateral model of migration. Once we take into account spatial dependence in both the dependent and the explanatory variables, specification tests reveal that our model shows no remaining unobserved heterogeneity.

Our evidence suggests that SSA migration results from a multidimensional set of factors. The results seem to confirm the hypothesis of Ratha and Shaw (2007) that South-South migration is to a large extent driven by income differences, networks and geographical proximity. On the other hand, we also find support for the role played by conflicts in the home country and relative freedom in the host country. Furthermore, deteriorating environmental conditions in a specific country discourage migration towards it. While for the economic determinants and migrant networks, the direct effects seem to dominate, our results suggest the presence of spillover effects (and hence a regional dimension) for the sociopolitical and environmental determinants.

As such, our results are in line with the main findings of the literature on South-South migration determinants, as discussed for instance in Bakewell (2009), for which we provide empirical evidence. Caution in generalizing these results to other contexts of South-South migration remains necessary, as the South combines a largely heterogeneous mixture of countries with idiosyncratic profiles and region specific developments. Yet, it should be clear that an analysis of migration in a South-South context should include economic determinants as well as other determinants that match the specificities of the

[17]The test statistics for destination, origin and combined effects, were 52.13, 36.50 and 16.91, respectively.

particular setting.

7 Acknowledgements

We wish to thank Paul Elhorst, Gerdie Everaert, Frédéric Docquier, Hubert Jayet, Tom Verbeke and two anonymous referees for their many valuable comments. For their valuable suggestions, we are also grateful to participants at the Fifth IZA/World Bank Conference Employment and Development (Cape Town, 2010), the Migration: a world in motion. A multinational conference on migration and migration policy (Maastricht, 2010), the IAB/HWWI Transnationality of migrants workshop (Bamberg, 2010), the 52nd Annual conference of the Indian Society of Labour Economics (Dharwad, 2010), the 7th World Conference of the Spatial Econometrics Association (Washington, 2013) and participants at internal seminars at Ghent University.

References

Adebusoye, P. (2006). Geographic labour mobility in Sub-Saharan Africa. *International Development Research Centre (IDRC) Working Paper*, 1.

Adepoju, A. (1995). Emigration dynamics in Sub-Saharan Africa. *International migration*, 33(3-4):313–390.

Agesa, J. and Agesa, R. U. (1999). Gender differences in the incidence of rural to urban migration: Evidence from kenya. *Journal of Development Studies*, 35(6):36–58.

Andersson, J. A. (2001). Reinterpreting the rural-urban connection: Migration practices and socio-cultural dispositions of Buhera workers in Harare. *Journal of the International African Institute*, 71(1):82–112.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion.* Princeton University Press.

Anselin, L. and Le Gallo, J. (2006). Interpolation of air quality measures in hedonic house price models: spatial aspects. *Spatial Economic Analysis*, 1(1):31–52.

Bakewell, O. (2009). South-South migration and human development: Reflections on African experiences. *Human Development Research Paper (HDRP) Series*, 2009/07.

Barkley, A. and McMillan, J. (1994). Political freedom and the response to economic incentives: Labor migration in Africa, 1972-1987. *Journal of development economics*, 45(2):393–406.

Barrios, S., Bertinelli, L., and Strobl, E. (2006). Climatic change and rural-urban migration: The case of sub-Saharan Africa. *Journal of Urban Economics*, 60(3):357–371.

Beine, M., Docquier, F., and Özden, C. (2011). Diasporas. *Journal of Development Economics*, 95(1):30–41.

Borjas, G. (1989). Economic theory and international migration. *International Migration Review*, 23(3):457–485.

Burke, M., Miguel, E., Satyanath, S., Dykema, J., and Lobell, D. (2009). Warming increases the risk of civil war in Africa. *Proceedings of the National Academy of Sciences*, 106(49):20670–20674.

Caudill, S. B. (1988). Practitioners' corner: An advantage of the linear probability model over probit or logit. *Oxford Bulletin of Economics and Statistics*, 50(4):425–427.

Curry, L. (1972). A spatial analysis of gravity flows. *Regional Studies*, 6(2):131–147.

Cushing, B. and Poot, J. (2004). Crossing boundaries and borders: Regional science advances in migration modelling. *Papers in Regional Science*, 83(1):317–338.

de Haan, A., Brock, K., and Coulibaly, N. (2002). Migration, livelihoods and institutions: Contrasting patterns of migration in Mali. *The Journal of Development Studies*, 38(5):37–58.

Dillon, A., Mueller, V., and Salau, S. (2011). Migratory responses to agricultural risk in Northern Nigeria. *American Journal of Agricultural Economics*, 93(4):1048–1061.

Dunlevy, J. (1980). Nineteenth-century European immigration to the United States: Intended versus lifetime settlement patterns. *Economic Development and Cultural Change*, 29(1):77–90.

Elhorst, J. (2010). *Handbook of Applied Spatial Analysis*, chapter Spatial panel data models, pages 377–407. Springer.

Elhorst, J. (2013). Matlab code for spatial panels. *Journal of Regional Science*, Forthcoming.

Ezra, M. and Kiros, G. (2001). Rural out-migration in the drought prone areas of Ethiopia: A multilevel analysis. *International Migration Review*, 35(3):749–771.

Fertig, M. (2001). The economic impact of EU-enlargement: Assessing the migration potential. *Empirical Economics*, 26(4):707–720.

Findley, S. (1994). Does drought increase migration? A study of migration from rural Mali during the 1983-1985 drought. *International Migration Review*, 28(3):539–553.

Gallardo-Sejas, H., Gil-Pareja, S., Llorca-Vivero, R., and Martínez-Serrano, J. (2006). Determinants of European immigration: A cross-country analysis. *Applied Economics Letters*, 13(12):769–773.

Griffith, D. (2007). Spatial structure and spatial interaction: 25 years later. *The Review of Regional Studies*, 37(1):28–38.

Hampshire, K. (2002). Fulani on the move: Seasonal economic migration in the Sahel as a social process. *Journal of Development Studies*, 38(5):15–36.

Harris, J. and Todaro, M. (1970). Migration, unemployment and development: A two-sector analysis. *The American Economic Review*, 60(1):126–142.

Hatton, T. (1995). A model of UK migration, 1870-1913. *Review of Economics and Statistics*, 77(3):407–415.

Hatton, T. and Williamson, J. (2002). What fundamentals drive world migration? *NBER Working Paper*, 9159.

Haug, R. (2002). Forced migration, processes of return and livelihood construction among pastoralists in northern Sudan. *Disasters*, 26(1):70–84.

Hooghe, M., Trappers, A., Meuleman, B., and Reeskens, T. (2008). Migration to European countries: A structural explanation of patterns, 1980-2004. *International Migration Review*, 42(2):476–504.

Hughes, D. (1999). Refugees and squatters: Immigration and the politics of territory on the Zimbabwe-Mozambique border. *Journal of Southern African Studies*, 25(4):533–552.

Karemera, D., Oguledo, V. I., and Davis, B. (2000). A gravity model analysis of international migration to North America. *Applied Economics*, 32(13):1745–1755.

Kelejian, H., Tavlas, G., and Hondroyiannis, G. (2006). A spatial modelling approach to contagion among emerging economies. *Open economies review*, 17(4):423–441.

Koser, K. (1997). Information and repatriation: The case of Mozambican refugees in Malawi. *Journal of refugee studies*, 10(1):1–17.

LeSage, J. and Pace, R. (2008). Spatial econometric modeling of origin-destination flows. *Journal of Regional Science*, 48(5):941–967.

LeSage, J. and Pace, R. (2009). *Introduction to Spatial Econometrics*. CRC Press.

Lewer, J. and Van den Berg, H. (2008). A gravity model of immigration. *Economics letters*, 99(1):164–167.

Lucas, R. (1985). Mines and migrants in South Africa. *The American Economic Review*, pages 1094–1108.

Lucas, R. (1987). Emigration to South Africa's mines. *The American Economic Review*, pages 313–330.

Marchiori, L., Maystadt, J., and Schumacher, I. (2012). The impact of weather anomalies on migration in sub-Saharan Africa. *Journal of Environmental Economics and Management*.

Mayda, A. (2010). International migration: A panel data analysis of the determinants of bilateral flows. *Journal of Population Economics*, 23:1249–1274.

Meze-Hausken, E. (2000). Migration caused by climate change: How vulnerable are people inn dryland areas? *Mitigation and Adaptation Strategies for Global Change*, 5(4):379–406.

Mitze, T. (2009). Network dependency in migration flows: A space-time analysis for Germany since re-unification. *RUHR Economic Papers*, 205.

Ncube, M., Lufumpa, C., and Ndikumana, L. (2010). Africans abroad: migration trends, patterns and determinants. *Economic Brief*, 1(1):1–7.

Ortega, F. and Peri, G. (2009). The causes and effects of international migrations: Evidence from OECD countries 1980-2005. *NBER Working Paper Series*, w14833.

Ortega, F. and Peri, G. (2011). The aggregate effects of trade and migration: Evidence from OECD countries.

Pace, R. and LeSage, J. (2006). Interpreting spatial econometric models. In *North American Meeting of the Regional Science Association International, Toronto, CA*.

Pedersen, P. J., Pytlikova, M., and Smith, N. (2008). Selection and network effects - Migration flows into OECD countries 1990-2000. *European Economic Review*, 52(7):1160–1186.

Ratha, D. and Shaw, W. (2007). South-South migration and remittances. *World Bank Working Paper*, 102.

Ruyssen, I., Everaert, G., and Rayp, G. (2012). Determinants and dynamics of migration to OECD countries in a three-dimensional panel framework. *Empirical Economics*.

Sen, A. and Smith, T. (1995). *Gravity models of spatial interaction behavior*. Advances in spatial and network economics. Springer-Verlag.

Sjaastad, L. A. (1962). The costs and returns of human migration. *The Journal of Political Economy*, 70(5):80–93.

Taylor, J. (1990). The reorganization of mine labor recruitment in Southern Africa: Evidence from Botswana. *International Migration Review*, 24(2):250–272.

Todaro, M. (1969). Model of labor migration and urban unemployment in less developed countries. *American Economic Review*, 59(1):138–148.

Verbeek, M. (2012). *A Guide to Modern Econometrics*. Wiley.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25.

Özden, C., Parsons, C., Schiff, M., and Walmsley, T. (2011). Where on earth is everybody? The evolution of global bilateral migration 1960-2000. *The World Bank Economic Review*, 25(1):12–56.

8 Appendix

8.1 Data sources

Migration data

- Migrant stocks ($\ln MST_{dot-1}$): the number of foreign residents in each destination in 1970 and 1980, disaggregated by country of origin. To avoid taking the log of zero, we add unity to each observation. *Source*: World Bank GBMD.

- Migrant rates (M_{dot}/N_{dt}): proxied by the change in MST_{dot} between 1980-1990 and 1990-2000 per 1000 of the average destination country's population. *Source*: World Bank GBMD and US Census Bureau's Population Estimates.

Explanatory variables

- Incomes ($\ln inc$): due to the lack of real wage data, average incomes are approximated by the log of gross domestic product per capita in purchasing power parities at 2005 constant prices. *Source*: Penn World Tables 7.0.

- Employment rates ($\ln empl$): log of the ratio of employed persons to the entire population. *Source*: compiled from the ILO's Key Indicators of the Labor Market, the Total Economy Database and the UN's Labor Force Statistics.

- Education ($\ln educ$): log of enrollment in secondary education divided by the population of the age group that typically corresponds to this level of education. *Source*: UNESCO Institute for Statistics.

- Share of the young population ($\ln youngpop$): log of the population aged between 0 and 14 as a percentage of the total population. *Source*: Africa Development Indicators (2010).

- Conflict ($\ln confl$): dichotomous variable capturing whether multiple regional wars took place during the decade. *Source*: Africa Migration Project's Violence and Unrest Variables.

- Disaster ($\ln disaster$): log of the share of the population affected (injured and deaths) by disasters such as droughts, earthquakes, epidemics, etc. (decade totals). *Source*: Emergency Events Database.

- Climate ($\ln clim$): log of temperature deviations from the century average (decade averages). *Source*: Intergovernmental Panel on Climate Change.

- Relative freedom ($\ln fr$): categorical variable which takes the values free, partly free and not free and reflects a combination of measures on civil liberties and political rights (decennium averages). In particular, political rights represent the degree of implementation or non-implementation of a country's democratic processes. Civil liberties reflect civil rights and desires in education, freedom of religion and choice of residence. *Source*: Freedom House.

- Distance ($\ln dist$): log of distance between the main cities (in population terms) of origin and destination countries. *Source*: CEPII Distance Database (2010).

- Contiguity (*commbord*), colonial ties (*commcol*) and common language (*commlang*): dichotomous variables coded 1 if origin and destination countries share respectively a common border, a former colonizer, or a common ethnological language (a language that is spoken by at least 9 percent of the population in both countries) and 0 otherwise. *Source*: CEPII Distance Database (2010).

- Regional economic integration (*regint*): dichotomous variable coded 1 if both countries were or became a member of the same regional economic community during the decade under consideration, and 0 otherwise. The regional economic communities taken into account are ECOWAS, ECCAS, IGAD and SADC.

Table 4: Migration stocks and changes by destination

	Migrant stocks			Migrant stock change		Migrant stock change/ Population$_d$*1000	
Destination	1980	1990	2000	1980-1990	1990-2000	1980-1990	1990-2000
Angola	7673	8426	13272	753	4846	0.104	0.511
Benin	55397	72877	125159	17480	52282	5.055	11.111
Botswana	1553	16547	41429	14994	24882	16.651	19.676
Burkina Faso	87135	153914	149175	66779	-4739	10.569	-0.567
Burundi	78283	64388	51640	-13895	-12748	-3.233	-2.303
Cameroon	145513	197749	166486	52236	-31263	5.961	-2.631
Cape Verde	5546	3283	3815	-2263	532	-7.634	1.566
Central African Republic	46327	54163	16943	7836	-37220	3.336	-12.065
Chad	54194	59896	67871	5702	7975	1.261	1.365
Comoros	13902	12073	11825	-1829	-248	-5.383	-0.576
Congo, Democratic Republic	182604	126189	103313	-56415	-22876	-1.945	-0.584
Congo, Republic	68643	113912	21023	45269	-92889	27.038	-40.992
Equatorial Guinea	3341	1736	2866	-1605	1130	-6.269	3.045
Ethiopia	143927	151311	159060	7384	7749	0.205	0.16
Gabon	63476	116867	177840	53391	60973	74.795	64.997
Gambia	71963	112565	162529	40602	49964	62.229	52.537
Ghana	55525	96175	136824	40650	40649	3.692	2.638
Guinea	18662	75214	238929	56552	163715	12.716	26.758
Guinea-Bissau	12507	11546	11094	-961	-452	-1.218	-0.454
Kenya	91953	100050	489530	8097	389480	0.496	16.672
Lesotho	3597	3066	3924	-531	858	-0.391	0.504
Liberia	69842	69842	66437	0	-3405	0	-1.592
Madagascar	3155	9163	8278	6008	-885	0.691	-0.076
Malawi	283745	278751	273844	-4994	-4907	-0.798	-0.514
Mali	99705	97873	78225	-1832	-19648	-0.269	-2.36
Mauritania	29193	50284	55570	21091	5286	13.652	2.746
Mauritius	1671	1082	2680	-589	1598	-0.611	1.505
Mozambique	11214	79249	294579	68035	215330	5.622	16.578
Namibia	61072	97899	93733	36827	-4166	34.807	-2.833
Niger	70811	112483	156589	41672	44106	6.839	5.625
Nigeria	1010988	367636	616421	-643352	248785	-8.598	2.573
Rwanda	38447	43482	346505	5035	303023	0.98	43.293
Senegal	96860	180641	178114	83781	-2527	14.931	-0.344
Sierra Leone	87765	86387	87498	-1378	1111	-0.413	0.263
Somalia	10721	15285	15688	4564	403	0.788	0.06
South Africa	534447	916630	708724	382183	-207906	13.065	-5.403
Swaziland	27998	30004	33841	2006	3837	3.281	4.352
Tanzania	381619	295308	229336	-86311	-65972	-4.624	-2.616
Togo	132878	138798	143000	5920	4202	2.255	1.129
Uganda	477926	315123	240689	-162803	-74434	-13.114	-4.264
Zambia	195614	129824	117469	-65790	-12355	-11.659	-1.572
Zimbabwe	395848	373347	343378	-22501	-29969	-3.138	-2.951

Table 5: Migration stocks and changes by origin

	Migrant stocks			Migrant stock change		Migrant stock change/ Population$_d$*1000	
Destination	1980	1990	2000	1980-1990	1990-2000	1980-1990	1990-2000
Angola	80893	106562	109996	25669	3434	3.562	0.362
Benin	338087	206629	283809	-131458	77180	-38.015	16.402
Botswana	77211	71831	33345	-5380	-38486	-5.975	-30.433
Burkina Faso	98499	103114	120524	4615	17410	0.73	2.082
Burundi	266845	176588	149908	-90257	-26680	-20.999	-4.82
Cameroon	107246	74011	91552	-33235	17541	-3.793	1.476
Cape Verde	12374	12689	29803	315	17114	1.063	50.371
Central African Republic	30319	23298	17078	-7021	-6220	-2.989	-2.016
Chad	105600	139195	111939	33595	-27256	7.429	-4.666
Comoros	5883	13099	12981	7216	-118	21.239	-0.274
Congo, Democratic Republic	227587	244635	493524	17048	248889	0.588	6.357
Congo, Republic	24900	27005	39873	2105	12868	1.257	5.679
Equatorial Guinea	21854	31832	45731	9978	13899	38.975	37.454
Ethiopia	7061	8957	22895	1896	13938	0.053	0.288
Gabon	4014	4554	10198	540	5644	0.756	6.017
Gambia	16881	16859	16466	-22	-393	-0.034	-0.413
Ghana	276337	177951	224906	-98386	46955	-8.935	3.047
Guinea	219223	270879	251557	51656	-19322	11.615	-3.158
Guinea-Bissau	42044	59574	64005	17530	4431	22.221	4.449
Kenya	159459	96745	99833	-62714	3088	-3.84	0.132
Lesotho	215510	324547	171044	109037	-153503	80.258	-90.123
Liberia	29077	54128	157105	25051	102977	13.489	48.152
Madagascar	17350	15665	16030	-1685	365	-0.194	0.031
Malawi	249004	255780	213695	6776	-42085	1.083	-4.409
Mali	259250	209038	302577	-50212	93539	-7.361	11.233
Mauritania	55426	69903	75008	14477	5105	9.371	2.652
Mauritius	11131	14341	10187	3210	-4154	3.331	-3.913
Mozambique	395272	458518	565895	63246	107377	5.226	8.267
Namibia	64565	104499	57694	39934	-46805	37.744	-31.828
Niger	156895	99767	153531	-57128	53764	-9.375	6.856
Nigeria	211522	221861	231364	10339	9503	0.138	0.098
Rwanda	333590	277895	162916	-55695	-114979	-10.836	-16.427
Senegal	126078	166829	201451	40751	34622	7.262	4.712
Sierra Leone	22460	45179	115148	22719	69969	6.812	16.548
Somalia	100049	103345	144118	3296	40773	0.569	6.093
South Africa	107432	146279	264060	38847	117781	1.328	3.061
Swaziland	45525	71912	44058	26387	-27854	43.154	-31.595
Tanzania	143025	123466	153438	-19559	29972	-1.048	1.189
Togo	182107	130980	185378	-51127	54398	-19.473	14.62
Uganda	79510	78675	393485	-835	314810	-0.067	18.035
Zambia	115442	142056	121640	26614	-20416	4.716	-2.598
Zimbabwe	190703	260368	275400	69665	15032	9.716	1.48

Table 6: Summary statistics

Variable	Obs	Mean	S.D.	Min	Max
M_{dot}/N_{dt}	3444	0.130	1.670	-26.868	42.879
$\ln inc_{dt}$	3444	6.985	0.785	5.445	9.477
$\ln inc_{ot}$	3444	3.893	0.288	2.989	4.481
$\ln empl_{dt}$	3444	3.232	0.749	1.389	4.671
$\ln empl_{ot}$	3444	3.809	0.071	3.423	3.909
$\ln MST_{dot-1}$	3444	3.203	3.211	0.000	12.566
$\ln educs_{dt}$	3444	0.080	0.196	0.000	0.693
$\ln educs_{ot}$	3444	0.297	0.361	0.000	1.099
$\ln youngpop_{dt}$	3444	0.252	0.336	0.000	1.553
$\ln youngpop_{ot}$	3444	1.157	0.237	0.713	1.493
$\ln confl_{dt}$	3444	6.985	0.785	5.445	9.477
$\ln confl_{ot}$	3444	3.893	0.288	2.989	4.481
$\ln fr_{dt}$	3444	3.232	0.749	1.389	4.671
$\ln fr_{ot}$	3444	3.809	0.071	3.423	3.909
$\ln disaster_{dt}$	3444	0.297	0.361	0.000	1.099
$\ln disaster_{ot}$	3444	0.751	0.314	0.000	1.099
$\ln climate_{dt}$	3444	0.252	0.336	0.000	1.553
$\ln climate_{ot}$	3444	1.157	0.237	0.713	1.493
$\ln distance_{do}$	3444	7.925	0.760	2.349	9.178
$commbord_{do}$	3444	0.085	0.279	0.000	1.000
$commcol_{do}$	3444	0.254	0.436	0.000	1.000
$commlang_{do}$	3444	0.310	0.463	0.000	1.000
$regint_{do}$	3444	0.230	0.424	0.000	2.000
$W_d \ln inc_{dt}$	3444	6.549	1.879	0.000	8.717
$W_o \ln inc_{ot}$	3444	3.617	1.016	0.000	4.194
$W_d \ln empl_{dt}$	3444	2.999	0.937	0.000	4.671
$W_o \ln empl_{ot}$	3444	3.544	0.984	0.000	3.874
$W_w \ln MST_{dot-1}$	3444	3.119	1.450	0.000	12.002
$W_d \ln educs_{dt}$	3444	0.077	0.137	0.000	0.693
$W_o \ln educs_{ot}$	3444	0.272	0.200	0.000	0.749
$W_d \ln youngpop_{dt}$	3444	0.221	0.204	0.000	0.914
$W_o \ln youngpop_{ot}$	3444	1.070	0.340	0.000	1.425
$W_d \ln confl_{dt}$	3444	6.484	1.845	0.000	9.477
$W_o \ln confl_{ot}$	3444	3.615	1.021	0.000	4.481
$W_d \ln fr_{dt}$	3444	3.001	0.923	0.000	4.671
$W_o \ln fr_{ot}$	3444	3.537	0.982	0.000	3.909
$W_d \ln disaster_{dt}$	3444	0.276	0.208	0.000	1.099
$W_o \ln disaster_{ot}$	3444	0.699	0.262	0.000	1.099
$W_d \ln climate_{dt}$	3444	0.234	0.193	0.000	1.553
$W_o \ln climate_{ot}$	3444	1.075	0.324	0.000	1.493
$W_w \ln distance_{do}$	3444	7.315	2.105	0.000	8.991
$W_w commbord_{do}$	3444	0.094	0.104	0.000	1.000
$W_w commcol_{do}$	3444	0.244	0.165	0.000	1.000
$W_w commlang_{do}$	3444	0.306	0.188	0.000	1.000
$W_w regint_{do}$	3444	0.224	0.149	0.000	1.000

Note: The sample includes 42 destination and origin countries.

Table 7: Correlation coefficients

		1	2	3	4	5	6	7	8	9	10	11
1	$\ln inc_{dt}$	1.000										
2	$\ln inc_{ot}$	−0.024	1.000									
3	$\ln empl_{dt}$	−0.314***	0.003	1.000								
4	$\ln empl_{ot}$	0.003	−0.314***	0.225***	1.000							
5	$\ln MST_{dot-1}$	−0.023	−0.088***	0.058***	0.069***	1.000						
6	$\ln educs_{dt}$	0.598***	−0.015	−0.238***	0.042**	0.026	1.000					
7	$\ln educs_{ot}$	−0.015	0.598***	0.042**	−0.238***	−0.105***	−0.019	1.000				
8	$\ln youngpop_{dt}$	−0.376***	0.009	0.127***	0.016	0.042**	−0.173***	0.007	1.000			
9	$\ln youngpop_{ot}$	0.009	−0.376***	0.016	0.127***	0.081***	0.007	−0.173***	−0.023	1.000		
10	$\ln confl_{dt}$	−0.135***	0.004	0.186***	−0.016	0.022	−0.227***	0.004	0.041**	−0.002	1.000	
11	$\ln confl_{ot}$	0.004	−0.135***	−0.016	0.186***	−0.031*	0.004	−0.227***	−0.002	0.041**	−0.024	1.000
12	$\ln fr_{dt}$	0.340***	−0.008	−0.144***	−0.014	−0.027	0.448***	−0.013	−0.172***	0.003	−0.165***	0.005
13	$\ln fr_{ot}$	−0.008	0.340***	−0.014	−0.144***	−0.101***	−0.013	0.448***	0.003	−0.172***	0.005	−0.165***
14	$\ln disaster_{dt}$	−0.262***	0.006	0.107***	0.032*	0.016	−0.247***	0.011	0.333***	−0.006	0.199***	−0.006
15	$\ln disaster_{ot}$	0.006	−0.262***	0.032*	0.107***	0.013	0.011	−0.247***	−0.006	0.333***	−0.006	0.199***
16	$\ln climate_{dt}$	−0.007	0.000	0.333***	−0.002	0.024	−0.010	0.001	−0.276***	0.007	0.082***	−0.002
17	$\ln climate_{ot}$	0.000	−0.007	−0.002	0.333***	−0.012	0.001	−0.010	0.007	−0.276***	−0.002	0.082***
18	$\ln distance_{do}$	0.021	0.021	−0.083***	−0.083***	−0.553***	0.000	0.000	−0.102***	−0.102***	0.016	0.016
19	$commbord_{do}$	0.015	0.015	−0.015	−0.015	0.214***	0.011	0.011	0.011	0.011	−0.118***	−0.118***
20	$commcol_{do}$	0.111***	0.111***	0.048***	0.048***	0.169***	0.157***	0.157***	−0.069***	−0.069***	−0.126***	−0.126***
21	$commlang_{do}$	−0.010	−0.010	0.027	0.027	0.557***	−0.036**	−0.036**	0.043**	0.043**	0.008	0.008
22	$regint_{do}$	0.049***	0.049***	−0.027	−0.027	0.467***	0.038**	0.038**	−0.040**	−0.040**	−0.004	−0.004

		12	13	14	15	16	17	18	19	20	21	22
12	$\ln fr_{dt}$	1.000										
13	$\ln fr_{ot}$	−0.023	1.000									
14	$\ln disaster_{dt}$	0.150***	−0.006	1.000								
15	$\ln disaster_{ot}$	−0.006	0.150***	−0.020	1.000							
16	$\ln climate_{dt}$	−0.168***	0.004	−0.354***	0.009	1.000						
17	$\ln climate_{ot}$	0.004	−0.168***	0.009	−0.354***	−0.024	1.000					
18	$\ln distance_{do}$	0.126***	0.126***	0.079***	0.079***	−0.046***	−0.046***	1.000				
19	$commbord_{do}$	0.067***	0.067***	0.063***	0.063***	−0.041*	−0.041**	−0.097***	1.000			
20	$commcol_{do}$	0.155***	0.155***	−0.056***	−0.056***	0.075***	0.075***	−0.085***	0.410***	1.000		
21	$commlang_{do}$	−0.033*	−0.033*	−0.003	−0.003	−0.016	−0.016	−0.387***	0.138***	0.130***	1.000	
22	$regint_{do}$	0.045***	0.045***	0.009	0.009	−0.055***	−0.055***	−0.516***	0.092***	0.045***	0.366***	1.000

Note: *, ** and *** indicate significance at the 10%, 5% and 1% level, respectively.

Table 8: Spatial Durbin model estimates

Dependent variable: $\ln M_{dot} / \ln N_{dt}$ — Sample period: 1980-2000

Variable	I	II	III	IV	V	VI	VII	VIII	IX
$\ln inc_{dt}$	0.101^{***}		0.163^{***}	0.156^{***}	0.202^{***}	0.213^{***}	0.240^{***}	0.256^{***}	0.243^{***}
	(2.804)		(3.142)	(3.074)	(3.619)	(3.582)	(3.212)	(3.410)	(3.329)
$\ln inc_{ot}$	-0.093^{***}		-0.067^{**}	−0.043	−0.013	−0.001	0.004	0.000	−0.012
	(−2.907)		(−2.074)	(−1.357)	(−0.358)	(−0.021)	(0.072)	(0.008)	(−0.257)
$\ln empl_{dt}$		−0.040	−0.005	−0.051	0.060	0.071	0.035	0.019	0.029
		(−0.448)	(−0.044)	(−0.462)	(0.534)	(0.558)	(0.336)	(0.185)	(0.239)
$\ln empl_{ot}$		0.007	-0.194^{**}	-0.218^{**}	−0.140	−0.130	-0.152^{*}	-0.170^{*}	−0.094
		(0.076)	(−2.052)	(−2.197)	(−1.375)	(−1.273)	(−1.671)	(−1.750)	(−0.731)
$\ln MST_{dot-1}$				0.069^{***}	0.018	0.023^{*}	0.024^{*}	0.028^{*}	0.032^{**}
				(3.266)	(1.606)	(1.736)	(1.704)	(1.949)	(2.158)
$\ln educs_{dt}$							−0.068	-0.083^{*}	-0.128^{**}
							(−1.401)	(−1.749)	(−1.977)
$\ln educs_{ot}$							−0.015	−0.022	−0.025
							(−0.316)	(−0.312)	(−0.382)
$\ln youngpop_{dt}$							−0.057	−0.088	0.071
							(−0.184)	(−0.304)	(0.283)
$\ln youngpop_{ot}$							0.145	0.225	−0.008
							(0.371)	(0.504)	(−0.030)
$\ln confl_{dt}$								0.025	0.238
								(0.182)	(1.457)
$\ln confl_{ot}$								0.178	0.189
								(1.106)	(1.189)
$\ln fr_{dt}$								−0.005	0.075
								(−0.031)	(0.384)
$\ln fr_{ot}$								0.095	0.120
								(0.634)	(0.810)
$\ln disaster_{dt}$									-0.033^{*}
									(−1.920)
$\ln disaster_{ot}$									−0.012
									(−1.580)
$\ln climate_{dt}$									0.108
									(0.264)
$\ln climate_{ot}$									−0.130
									(−0.838)
$\ln distance_{do}$					-0.149^{*}	-0.178^{*}	−0.187	−0.202	−0.191
					(−1.655)	(−1.870)	(−1.549)	(−1.496)	(−1.347)
$commbord_{do}$					0.600^{**}	0.614^{**}	0.591^{**}	0.565^{**}	0.546^{**}
					(2.227)	(2.296)	(2.213)	(2.131)	(2.054)
$commcol_{do}$					0.007	0.015	0.009	0.021	0.034
					(0.073)	(0.145)	(0.087)	(0.178)	(0.290)
$commlang_{do}$					0.067	0.053	0.073	0.059	0.037
					(0.777)	(0.602)	(0.765)	(0.743)	(0.496)
$regint_{do}$						−0.154	−0.155	−0.185	−0.182
						(−1.482)	(−1.463)	(−1.658)	(−1.530)
$W_d \ln inc_{dt}$	0.000		−0.045	−0.052	0.015	0.030	0.071	0.072	−0.005
	(0.046)		(−1.408)	(−1.408)	(0.220)	(0.430)	(0.689)	(0.610)	(−0.047)
$W_o \ln inc_{ot}$	0.006		−0.025	−0.011	0.019	0.024	0.023	0.001	−0.034
	(0.845)		(−0.465)	(−0.246)	(0.486)	(0.604)	(0.216)	(0.010)	(−0.275)

Continued on next page

Continued from previous page

Variable	I	II	III	IV	V	VI	VII	VIII	IX
$W_d \ln empl_{dt}$		0.023**	0.108*	0.036	0.338*	0.317	0.060	0.339	0.649**
		(2.266)	(1.694)	(0.537)	(1.683)	(1.563)	(0.270)	(1.503)	(2.059)
$W_o \ln empl_{ot}$		0.040***	0.077	0.052	−0.006	−0.015	0.208	0.151	0.033
		(2.826)	(0.772)	(0.588)	(−0.081)	(−0.193)	(1.296)	(0.968)	(0.199)
$W_w \ln MST_{dot-1}$				0.077	0.005	0.010	0.013	0.036	0.057
				(1.139)	(0.139)	(0.255)	(0.289)	(0.601)	(0.887)
$W_d \ln educs_{dt}$							−0.097	−0.155	−0.243
							(−0.816)	(−0.902)	(−1.243)
$W_o \ln educs_{ot}$							0.060	0.087	0.061
							(0.544)	(0.672)	(0.491)
$W_d \ln youngpop_{dt}$							0.319	0.042	−0.013
							(0.689)	(0.100)	(−0.027)
$W_o \ln youngpop_{ot}$							−0.281	−0.216	−0.171
							(−1.554)	(−1.160)	(−0.888)
$W_d \ln confl_{dt}$								−0.295	0.037
								(−0.934)	(0.119)
$W_o \ln confl_{ot}$								0.450	0.505
								(1.490)	(1.587)
$W_d \ln fr_{dt}$								0.555	0.769**
								(1.522)	(2.157)
$W_o \ln fr_{ot}$								0.021	0.144
								(0.165)	(0.959)
$W_d \ln disaster_{dt}$									−0.069***
									(−2.729)
$W_o \ln disaster_{ot}$									−0.020
									(−1.109)
$W_d \ln climate_{dt}$									−0.443
									(−0.861)
$W_o \ln climate_{ot}$									0.453
									(1.469)
$W_w \ln distance_{do}$					−0.191	−0.192	−0.213	−0.210	−0.206
					(−1.367)	(−1.339)	(−1.299)	(−1.266)	(−1.360)
$W_w commbord_{do}$					1.147	1.207	1.121	0.939	0.872
					(1.516)	(1.590)	(1.496)	(1.401)	(1.291)
$W_w commcol_{do}$					0.043	0.051	−0.006	−0.065	−0.051
					(0.130)	(0.155)	(−0.021)	(−0.243)	(−0.192)
$W_w commlang_{do}$					−0.089	−0.075	−0.065	−0.122	−0.212
					(−0.299)	(−0.250)	(−0.212)	(−0.375)	(−0.583)
$W_w regint_{do}$						−0.158	−0.167	−0.258	−0.285
						(−0.787)	(−0.771)	(−1.010)	(−1.170)
$W_d M_{dot}$	0.012***	0.016***	0.017***	0.017***	0.017***	0.017***	0.017***	0.007**	0.018***
	(3.259)	(4.106)	(4.734)	(4.498)	(4.678)	(4.497)	(4.398)	(1.983)	(4.587)
$W_o M_{dot}$	0.285***	0.239***	0.185***	0.182***	0.182***	0.181***	0.263***	0.233***	0.158***
	(7.387)	(6.473)	(4.893)	(4.746)	(4.827)	(4.743)	(7.161)	(6.242)	(4.348)
$W_w M_{dot}$	−0.017	0.009	0.025	0.026	0.025	0.026	0.010	−0.035	0.032
	(−0.747)	(0.371)	(1.098)	(1.109)	(1.082)	(1.109)	(0.435)	(−1.575)	(1.391)

Notes: T-statistics in parenthesis are calculated using simulated standard errors based on 1000 bootstrap samples. *, ** and *** indicate significance at the 10%, 5% and 1% level, respectively. Number of observations: 3444.

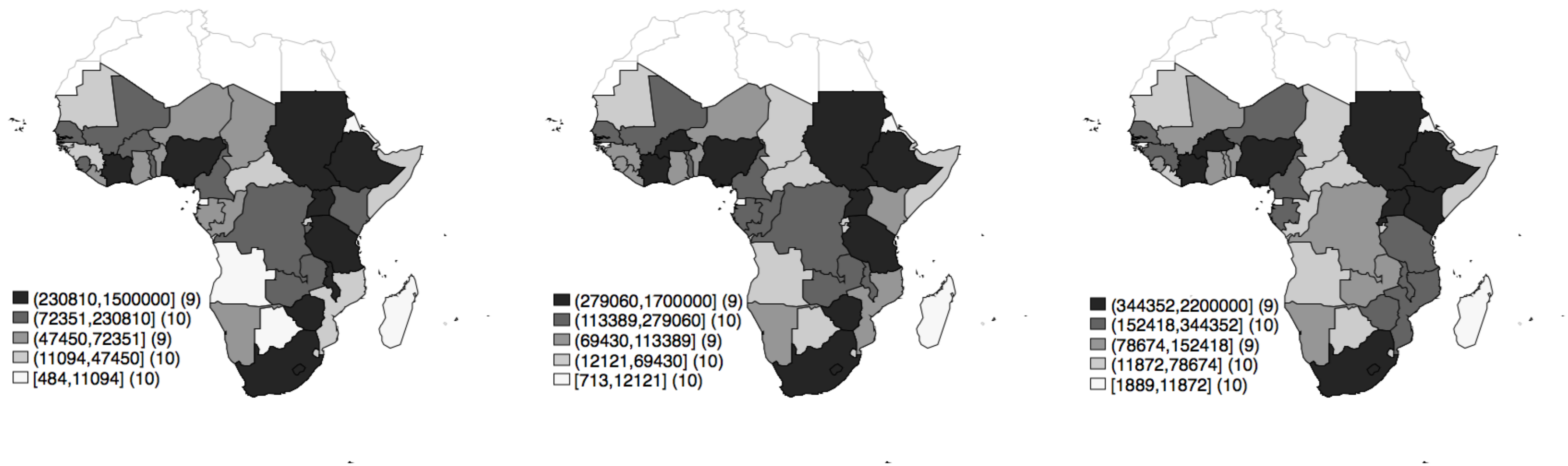
Table 9: Total effects estimates

Dependent variable: $\ln M_{dot}/\ln N_{dt}$									Sample period: 1980-2000
Variable	I	II	III	IV	V	VI	VII	VIII	IX
$\ln inc_{dt}$	0.103∗∗∗		0.120∗∗∗	0.103∗∗	0.221∗∗	0.246∗∗∗	0.320∗∗	0.323∗∗	0.243∗
	(3.020)		(2.837)	(2.036)	(2.422)	(2.691)	(2.259)	(2.077)	(1.672)
$\ln inc_{ot}$	−0.123∗∗∗		−0.111	−0.065	0.005	0.028	0.043	0.001	−0.052
	(−2.653)		(−1.380)	(−0.980)	(0.077)	(0.464)	(0.368)	(0.010)	(−0.411)
$\ln empl_{dt}$		−0.019	0.106	−0.017	0.413∗	0.392	0.107	0.354	0.704∗∗
		(−0.210)	(0.722)	(−0.115)	(1.733)	(1.564)	(0.423)	(1.399)	(1.982)
$\ln empl_{ot}$		0.062	−0.151	−0.202	−0.170	−0.177	0.070	−0.022	−0.075
		(0.526)	(−1.027)	(−1.370)	(−1.046)	(−1.079)	(0.353)	(−0.116)	(−0.411)
$\ln stock_{dot-1}$				0.152∗	0.028	0.034	0.036	0.060	0.097
				(1.804)	(0.700)	(0.762)	(0.803)	(1.029)	(1.395)
$\ln educs_{dt}$							−0.172	−0.236	−0.384
							(−1.269)	(−1.161)	(−1.570)
$\ln educs_{ot}$							0.056	0.087	0.043
							(0.433)	(0.654)	(0.359)
$\ln youngpop_{dt}$							0.237	−0.045	0.022
							(0.458)	(−0.084)	(0.037)
$\ln youngpop_{ot}$							−0.208	0.013	−0.199
							(−0.358)	(0.020)	(−0.582)
$\ln confl_{dt}$								−0.281	0.280
								(−0.777)	(0.822)
$\ln confl_{ot}$								0.837∗∗	0.838∗∗
								(2.179)	(2.224)
$\ln fr_{dt}$								0.562	0.870∗
								(1.097)	(1.657)
$\ln fr_{ot}$								0.152	0.321
								(0.691)	(1.347)
$\ln disaster_{dt}$									−0.105∗∗∗
									(−3.421)
$\ln disaster_{ot}$									−0.037∗
									(−1.730)
$\ln climate_{dt}$									−0.353∗
									(−1.659)
$\ln climate_{ot}$									0.388
									(1.044)
$\ln distance_{do}$					−0.355∗	−0.376∗∗	−0.388	−0.392∗	−0.406∗
					(−1.944)	(−1.963)	(−1.632)	(−1.733)	(−1.688)
$commbord_{do}$					1.856∗	1.816∗	1.745∗	1.492∗	1.497∗
					(1.899)	(1.907)	(1.851)	(1.772)	(1.705)
$commcol_{do}$					0.072	0.065	−0.002	−0.047	−0.017
					(0.243)	(0.221)	(−0.006)	(−0.223)	(−0.072)
$commlang_{do}$					−0.036	−0.019	0.009	−0.063	−0.190
					(−0.130)	(−0.069)	(0.034)	(−0.215)	(−0.557)
$regint_{do}$						−0.308	−0.308	−0.424	−0.484
						(−1.236)	(−1.184)	(−1.421)	(−1.539)

Notes: T-statistics in parenthesis are calculated using simulated standard errors based on 1000 bootstrap samples. *, ** and *** indicate significance at the 10%, 5% and 1% level, respectively. Number of observations: 3444.

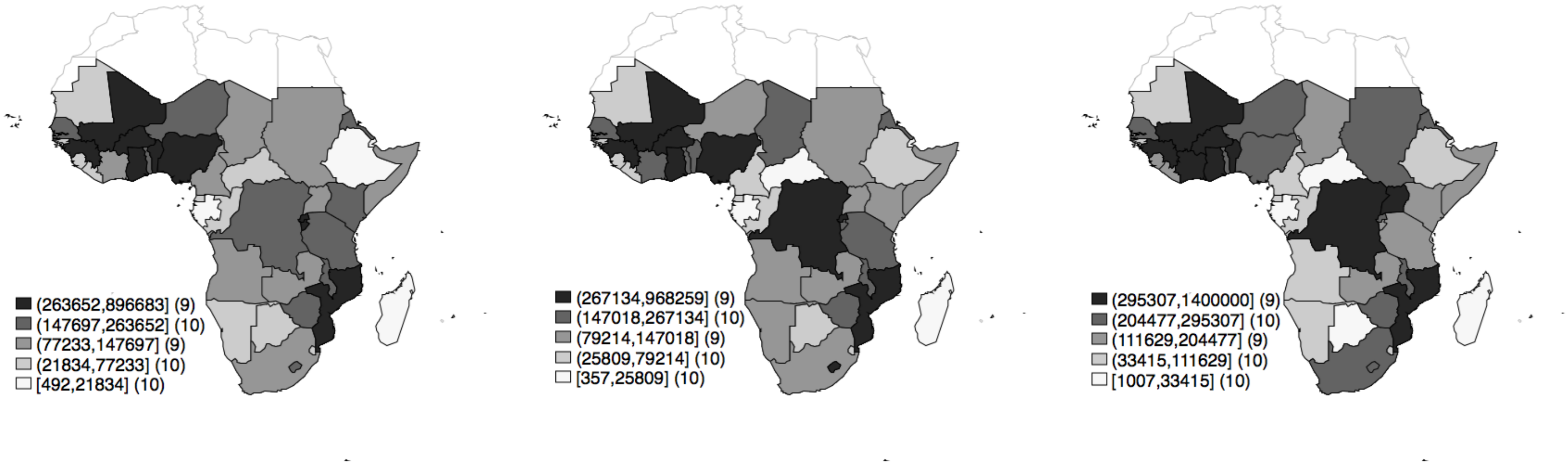
8.2 Figures

Figure 1: SSA migrant stocks by destination, 1980, 1990 and 2000



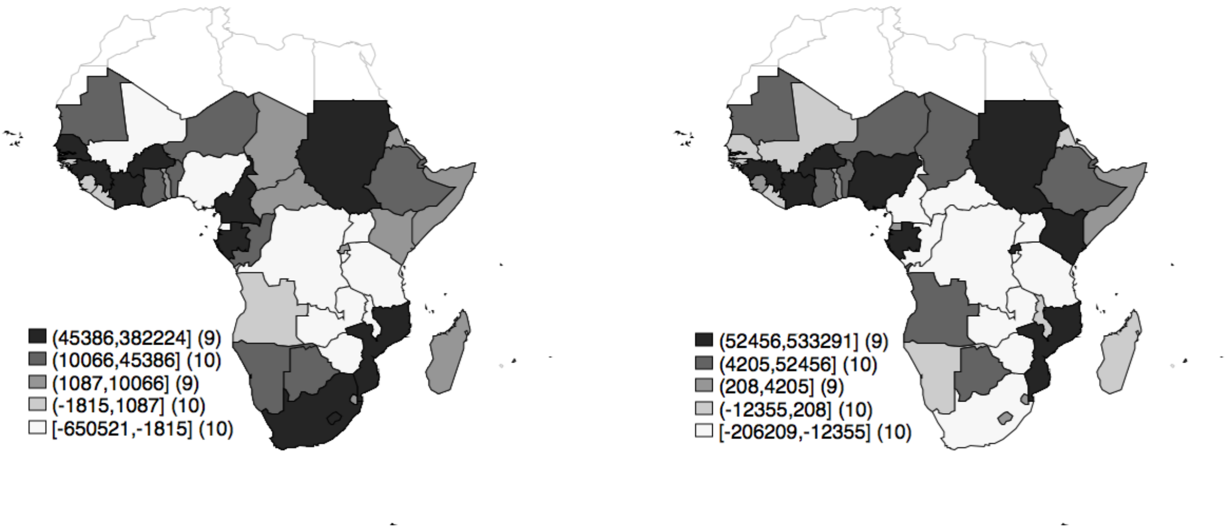
Source: Authors' calculations based on Global Bilateral Migration database, World Bank (2011)

Figure 2: SSA migrant stocks by origin, 1980, 1990 and 2000



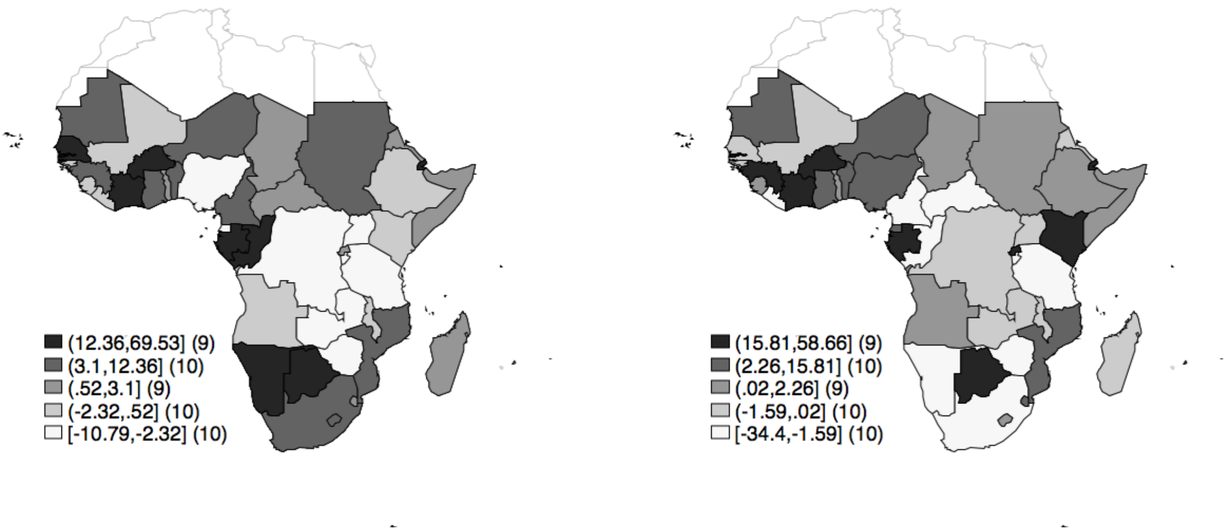
Source: Authors' calculations based on Global Bilateral Migration database, World Bank (2011)

Figure 3: SSA migrant stocks change by destination, 1980-1990 and 1990-2000



Source: Authors' calculations based on Global Bilateral Migration database, World Bank (2011)

Figure 4: SSA migrant stocks change by destination, 1980-1990 and 1990-2000 (population shares)



Source: Authors' calculations based on Global Bilateral Migration database, World Bank (2011)